DOCUMENT RESUME

ED 152 857 TM 007 128

AUTHOR Allison, Howard K., II
TITLE Rating Scale Format as it Affects Ratings in Student Teaching.
PUB DATE 29 Mar 78
NOTE 15p.; Paper presented at the Annual Meeting of the American Educational Research Association (62nd, Toronto, Ontario, Canada, March 27-31, 1978)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Analysis of Variance; Behavior Rating Scales; College Supervisors; *Comparative Testing; Cooperating Teachers; Elementary Secondary Education; *Error Patterns; *Evaluation Methods; Factor Analysis; Graphs; Higher Education; *Rating Scales; *Response Style (Tests); Scoring Formulas; *Student Teachers; Test Construction; Testing Problems; Test Reliability
IDENTIFIERS *Likert Scales

ABSTRACT

To assess the impact of various rating scale formats in rating student teachers, a Likert-type scale (Highly agree/slightly agree/neither/slightly disagree/highly disagree) was derived from the graphic scale currently used by the cooperating teachers and university supervisors. They rated the music student teachers on an eight-point scale from superior to average to inferior. Each format's effectiveness in dealing with the problems associated with rating scales was investigated, with the content of the items remaining constant. Data collected enabled the assessment of the contribution of each a leniency error, a halo effect, and a contrast error to ratings on each scale. By treating ratings as the dependent variable in a three way factorial analysis of variance, with rater, ratee, and item as the three independent variables, estimates of the three types of error are obtained in the first order interaction effects. The rating scale employing the graphic format yielded significant first order interactions representing each of the three types of constant error. The rating scale which employed the Likert-type format yielded only one significant first order interaction, that representing the leniency error. It appears that the Likert-type format was more effective in dealing with the halo effect and the contrast error. However, both formats appear to be susceptible to the leniency error. The 15-item grapic scale and the Likert-type derivation (52 items) are appended. (Author/CP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED152857

RATING SCALE FORMAT AS IT AFFECTS RATINGS IN STUDENT TEACHING

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Howard K. Allison II

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

by

Howard K. Allison II

Indiana University

TM007 128

A Paper Prepared for Presentation at the
Annual Meeting of the American Educational Research Association

Toronto, March 29, 1978

Printed in U. S. A.

PROBLEM

Problems associated with the use of rating scales have received considerable attention. Studies dealing with the effects of varying characteristics in rating scales have produced conflicting results (cf. Taylor and Hustman, 1956; Blumberg, DeSoto, and Kuethe, 1966; Froman, 1976; Taylor and Whitney, 1951; Madden and Bourdon, 1964; and Follman, Lucoff, Small, and Power, 1974). The present study attempts to examine the impact of varying rating scale format in rating student teachers. More specifically, the question was asked, does a contrived rating scale using a Likert-type format deal more effectively with problems associated with the use of rating scales than does a graphic format when content remains constant?

Problems associated with the use of rating scales are classified by Guilford (1954, 278-280) in terms of constant errors. Types of errors identified include the Error of Leniency which is a general constant tendency for a rater to rate too high or too low. The Halo Effect results in forcing the rating of an individual trait in the direction of the rater's general impression of the ratee. A Contrast Error is a source of bias in

which a rater has a tendency to rate others in the opposite direction from himself in a trait. The present study provided data which made possible the assessment of the impact of rating scale format upon these three types of constant error.

PROCEDURES

Subjects

Student teachers in the study were senior music education majors at a large mid-western university, assigned to student teaching situations representing all levels and areas of public school music education. Public school settings ranged from a large city and its metropolitan area to a medium-sized university town to small rural communities. Cooperating teachers in the study had all had prior experience in working with student teachers.

Materials

Two rating scales were used to collect data for the study. The first scale was a fifteen item graphic rating scale with eight response options. Each of the fifteen items is further defined by a number of statements which attempt to focus the rating for that item, but only one response is obtained for the item in general. This first scale is the rating scale presently used in evaluating student teaching at the university where the study was conducted. The second scale was derived from the first by making each of the defining statements of the original scale into a separate item. This resulted in 52 items, about half of which were phrased negatively, which were placed on a five-point Likert-type scale. This provided, for comparison, two scales with equal content but with a different format. A copy of each of the scales may be found in the appendix.

Data Collection

The study was conducted over a period of one semester and data employed from both mid-term and final evaluations. The cooperating teacher ratings, from both mid-term and final evaluations, were available as part of the data normally collected for each student teacher. One week before the date of the mid-term rating, each cooperating teacher was sent two copies of the derived form and a letter asking them to complete one copy of this second form, in addition to the original scale, at both the time of the mid-term and the final evaluation. The return rate on the derived form was 72%, providing complete data from cooperating teachers on 26 cases. Ratings on both scales were also obtained from university supervisors who made observational visits to student teachers.

Data Analysis

In examining the effect of rating scale format, the following score model is adopted (Guilford, 1959, 280-281):

$$X_{ijk} = X_{ijt} + X_{ijke}$$

where: X_{ijk} = a rating of person I in trait J by rater K

X_{ijt} = the true value of person I in trait J

X_{ijke} = the total error in rating X_{ijk}

The error component of the above model can be further partitioned into the following components:

X_{kl} = rater K's leniency error

X_{ki} = rater K's halo error in connection with person I

X_{kj} = rater-trait interaction error (e.g.: contrast error)

X_{ijkr} = residual error which includes all error variance not already identified

After partitioning the total error variance, the initial score model may be restated:

$$X_{ijk} = X_{ijt} + X_{kl} + X_{ki} + X_{kj} + X_{ijkr}$$

By treating ratings as the dependent variable in a three way factorial analysis of variance, with rater, ratee, and item as the three independent variables, estimates of the three types of constant error represented in the score model above are obtained in the first order interaction effects. (Guilford, 1954, 282) Since there is only one replicate per cell in such a design, and consequently no within cells variability, the second order interaction mean square was used as a conservative estimate of error variance in testing the significance of the first order interaction effects. (Dayton, 1970, 89-90) Treating the data in this manner, the rater by ratee interaction represents the halo effect. The rater by item interaction represents the contrast error. The ratee by item interaction represents the leniency error. Since not all student teachers were visited by a single university supervisor, these ratings were adjusted, using simple linear regression, on the basis of ratings given a common group of three student teachers who were not otherwise involved in the study. Adjustment of cooperating teacher ratings, on the basis of university supervisor ratings was also necessary in this phase of the data analysis in order to control for the fact that not all student teachers were rated by the same cooperating teacher.

In addition to this analysis of variance procedure used to assess constant error in each of the scales, several descriptive statistics were computed on the distribution from each scale. These include skewness and kurtosis in order to compare the shapes of the distributions, Pearson

product-moment correlation coefficients among all sets of ratings obtained, and an estimate of reliability for each scale.

RESULTS AND DISCUSSION

In examining the shape of the two distributions, the skewness of each (original scale -2.95; derived scale -1.88) indicates a negatively skewed distribution. An examination of the kurtosis, adjusted so that zero represents a normal distribution (original scale -1.46; derived scale -.86), reveals that each distribution is slightly more flat than a normal distribution. In both cases, however, the derived scale more closely approximates a normal distribution than does the original scale.

Cronbach's Alpha was computed as an estimate of scale reliability for each scale. Results in both cases were high, .93 for the original scale and .97 for the derived scale.

Table 1 presents the product-moment correlation coefficients which were computed among all sets of data collected. Aside from the correlation between the two forms when the rating was made by university supervisors, the highest correlation, .64, was found between ratings on the derived scale. It is interesting to note that cooperating teacher ratings between the two forms correlated only .39. All correlation coefficients were significant beyond the .05 level.

TABLE 1

Pearson Product-Moment Correlations Among Sets of Ratings

	Cooperating Teacher, Original Scale	Cooperating Teacher, Derived Scale	University Supervisor, Original Scale
Cooperating Teacher, Derived Scale	.39		
University Supervisor, Original Scale	.48	.63	
University Supervisor, Derived Scale	.46	.64	.67

Results of the analysis of variance performed on scores on the original scale is presented in Table 2. Examining the quasi F ratios testing the main effects, we find significant differences among items and ratees and no significant difference between raters, as would be expected. Examining the F ratios which conservatively test the first order interaction effects, we find each of the first order interaction effects to be significant. There was a significant difference attributable to the rater by ratee interaction, representing the halo effect. The significant interaction of rater and item

represents the contrast error. The significant ratee by item interaction represents the leniency error. Because of these significant interaction effects, we may conclude that there is systematic error variance included in the ratings obtained using the original scale representing each of the three types of constant error.

TABLE 2

Analysis of Variance Summary Table for Original Scale

Source	df	Sum of Squares	Mean Square	F	P
Rater	1	21.92	21.92	1.25	N.S.
Ratee	25	720.74	28.83	2.04	$< .05$
Item	14	298.81	21.20	1.89	$< .05$
Rater by Ratee	25	316.01	12.64	2.66	$< .05$
Rater by Item	14	135.94	9.71	2.04	$< .05$
Ratee by Item	350	2187.49	6.25	1.31	$< .01$
Rater by Ratee by Item	350	1665.98	4.76		

The results of the analysis of variance performed on scores on the derived scale is summarized in Table 3. Examining the quasi F ratios testing the main effects, again, as would be expected, we find significant differences among items and ratees and no significant difference between raters. Examining the F ratios which conservatively test the first order interaction

effects, we find the only significant difference in the ratee by item interaction, representing the leniency error. Since two of the first order interactions were not significant in a conservative test, the data was subjected to three separate two-way analyses of variance, pooling the within rater variance in one case, pooling the within ratee variance in the second case, and pooling the within item variance in the third case, which made possible a test of the two-way interactions using a more appropriate denominator. Again in this analysis, the only significant two-way interaction was ratee by item which represents the leniency error.

TABLE 3

Analysis of Variance Summary Table for Derived Scale

Source	df	Sum of Squares	Mean Square	F	P
Rater	1	18.66	18.66	3.68	N.S.
Ratee	25	985.11	39.40	7.74	< .01
Item	51	1756.92	34.45	6.93	< .01
Rater by Ratee	25	100.26	4.01	1.42	N.S.
Rater by Item	51	198.32	3.89	1.37	N.S.
Ratee by Item	1275	4974.43	3.91	1.38	< .01
Rater by Ratee by Item	1275	3609.54	2.83		

Table 4 provides a summary of the proportion of variance attributable to each of the variance components in both the original and the derived scales.

TABLE 4

Proportion of Variation Attributable to Specific Sources in Original and Derived Scales

Source of Variance	Proportion, Original Scale	Proportion, Derived Scale
Rater	.21	.17
Ratee	.27	.36
Item	.20	.32
Rater by Ratee (Halo Effect)	.12	.04
Rater by Item (Contrast Error)	.09	.04
Ratee by Item (Leniency Error)	.06	.04
Rater by Ratee by Item	.05	.03

Since with the original scale each of the first order interactions were significant and with the derived scale only the ratee by item interaction was significant, it appears that the Likert-type format employed in the derived scale was effective in eliminating error variance due to the halo effect and the contrast error. However, both the graphic type format of the original scale and the Likert-type format of the derived scale appear to be susceptible to the leniency error, as represented by the significant ratee by item interaction in both of the scales.

REFERENCES

Blumberg, Herbert H., DeSoto, Clinton B., and Kuethe, James L. "Evaluation of Rating Scale Formats." Personnel Psychology, 1966, XIX, 243-260.

Dayton, C. Mitchell. The Design of Educational Experiments. New York: McGraw-Hill, 1970.

Follman, John, Lucoff, Manny, Small, Leslie, and Power, Fred. "Kinds of Keys of Student Ratings of Faculty Teaching Effectiveness." Paper presented at the annual meeting of the American Educational Research Association, Chicago, Illinois, April, 1974.

Froman, Robin D. "The Influence of Format Change on the Halo Effect of Student Ratings." Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, California, April, 1976.

Guilford, J.P. Psychometric Methods. New York: McGraw-Hill, 1954.

Madden, Joseph M. and Bourdon, Roger D. "Effects of Variations in Rating Scale Format on Judgement." Journal of Applied Psychology, 1964, Vol. 48, No. 3, 147-151.

Taylor, Erwin K. and Hastman, Roy. "Relation of Format and Administration to Characteristics of Graphic Rating Scales." Personnel Psychology, 1956, IX, 181-206.

Taylor, Erwin K. and Wherry, Robert J. "A Study of Leniency in Two Rating Systems." Personnel Psychology, 1951, IV, 39-47.

APPENDIX

Final Evaluation of Student Teaching

Name of Student Teacher ..... Date .....

CHARACTERISTICS AFFECTING TEACHING SUCCESS

Directions: Check the appropriate position on each scale. Underline strong points. Encircle weak points.

	Superior	Average	Inferior
KNOWLEDGE OF SUBJECT MATTER			
1. Has an understanding and a working knowledge of content in teaching field.	| | | | | | | | |		
2. Is actively engaged in increasing his knowledge of subject matter.	| | | | | | | | |		
INSTRUCTIONAL CAPACITIES			
3. Ability to Organize Materials for Teaching Purposes: Makes adequate plans for teaching; selects materials with attention to individual differences; organizes materials effectively.	| | | | | | | | |		
4. Ability to Organize Learning Situations: Has general mastery of method; is able to create effective learning situations; obtains wide pupil participation; maintains proper balance between teacher-pupil activity; provides for individual differences.	| | | | | | | | |		
5. Class Achievement: Pupil learning continues at appropriate levels; selects appropriate appraisal techniques.	| | | | | | | | |		
6. Classroom Management: Is fair and just in dealing with pupils; understands the psychological, physiological, and sociological needs of his pupils; has the interest and cooperation of pupils; encourages social responsibility in pupils.	| | | | | | | | |		
PERSONAL/PROFESSIONAL QUALITIES			
7. Classroom Personality: Is mentally alert; has sense of humor; exercises self-control; displays drive and vitality.	| | | | | | | | |		
8. Social Qualities: Is friendly, understanding and helpful; has ability to get along with others and can empathize with their problems; is courteous and tactful; is interested in pupils.	| | | | | | | | |		
9. Professional Zeal: Is interested in teaching; takes steps toward self-improvement; is an enthusiastic worker; believes that teaching is worthwhile.	| | | | | | | | |		
10. General Knowledge and Information: Has a wide variety of interests and a broad understanding of social concerns.	| | | | | | | | |		
11. Loyalty and Cooperation: Is dependable and punctual; is willing and able to take suggestions and criticisms; cooperates with associates and supervisors; upholds school policies.	| | | | | | | | |		
12. Use of Oral and Written English: Presents ideas simply and clearly; uses good English in and out of the classroom.	| | | | | | | | |		
13. Health: Has good general physical condition and mental health; able to carry normal load with energy in reserve.	| | | | | | | | |		
14. Personal Appearance: Exhibits good taste and neatness in dress; is clean; has no distracting mannerisms.	| | | | | | | | |		
15. Voice and Speech: Is clear and distinct; has good inflection and modulation; is easy to understand; uses correct pronunciation; is free from irritating mannerisms.	| | | | | | | | |		

STUDENT TEACHER EVALUATION FORM

Name of Student Teacher _____

The purpose of the following statements is to help you as accurately as possible evaluate the student teacher named above. Respond to each statement on the basis of how much you agree or disagree that the statement is descriptive of the student teacher. Use the following five point scale:

HD - highly disagree that the statement is descriptive
D - slightly disagree that the statement is descriptive
NN - neither disagree nor agree that the statement is descriptive
A - slightly agree that the statement is descriptive
HA - highly agree that the statement is descriptive

Please choose only one response to each statement. Please attempt to answer every statement. CIRCLE RESPONSES

HD D NN A HA	1.	Has an understanding and working knowledge of content in her/his teaching field.
HD D NN A HA	2.	Is actively engaged in increasing her/his knowledge of subject matter.
HD D NN A HA	3.	Makes inadequate plans for teaching.
HD D NN A HA	4.	Selects materials with attention to individual differences.
HD D NN A HA	5.	Organizes materials effectively.
HD D NN A HA	6.	Has general mastery of methods.
HD D NN A HA	7.	Is not able to create effective learning situations.
HD D NN A HA	8.	Does not obtain wide pupil participation.
HD D NN A HA	9.	Maintains proper balance between teacher/pupil activity.
HD D NN A HA	10.	Does not provide for individual differences.
HD D NN A HA	11.	Does not provide for pupil learning to continue at appropriate levels.
HD D NN A HA	12.	Selects appropriate appraisal techniques.
HD D NN A HA	13.	Is fair and just in dealing with pupils.
HD D NN A HA	14.	Does not understand the psychological needs of her/his pupils.
HD D NN A HA	15.	Does not understand the physiological needs of her/his pupils.
HD D NN A HA	16.	Understands the sociological needs of her/his pupils.
HD D NN A HA	17.	Does not have the interest of pupils.
HD D NN A HA	18.	Has the cooperation of pupils.
HD D NN A HA	19.	Encourages social responsibility in pupils.
HD D NN A HA	20.	Is mentally alert.
HD D NN A HA	21.	Does not have a sense of humor.
HD D NN A HA	22.	Exercises self-control.
HD D NN A HA	23.	Does not display drive and vitality.
HD D NN A HA	24.	Is not friendly.
HD D NN A HA	25.	Is understanding.
HD D NN A HA	26.	Is not helpful.
HD D NN A HA	27.	Does not have the ability to get along with others.
HD D NN A HA	28.	Cannot empathize with the problems of others.
HD D NN A HA	29.	Is courteous.
HD D NN A HA	30.	Is not tactful.

HD - highly disagree that the statement is descriptive
D - slightly disagree that the statement is descriptive
NN - neither disagree nor agree that the statement is descriptive
A - slightly agree that the statement is descriptive
HA - highly agree that the statement is descriptive

HD D NN A HA	31.	Is interested in pupils.
HD D NN A HA	32.	Is not interested in teaching.
HD D NN A HA	33.	Takes steps toward self-improvement.
HD D NN A HA	34.	Is an enthusiastic worker.
HD D NN A HA	35.	Believes teaching is worthwhile.
HD D NN A HA	36.	Has narrow interests.
HD D NN A HA	37.	Has a broad understanding of social concerns.
HD D NN A HA	38.	Is dependable and punctual.
HD D NN A HA	39.	Is willing and able to take suggestions and criticisms.
HD D NN A HA	40.	Does not uphold school policies.
HD D NN A HA	41.	Is uncooperative with associates and supervisors.
HD D NN A HA	42.	Presents ideas simply and clearly.
HD D NN A HA	43.	Uses good English in and out of the classroom.
HD D NN A HA	44.	Does not have good general physical condition and mental health
HD D NN A HA	45.	Is not able to carry a normal load with energy in reserve.
HD D NN A HA	46.	Does not exibit good taste and neatness in dress.
HD D NN A HA	47.	Has no distracting physical mannerisms.
HD D NN A HA	48.	Her/his speech is clear and distinct.
HD D NN A HA	49.	Her/his speech is not characterized by good inflection and modulation.
HD D NN A HA	50.	Is not easy to understand.
HD D NN A HA	51.	Uses incorrect pronunciation.
HD D NN A HA	52.	Her/his speech is free from irritating mannerisms.